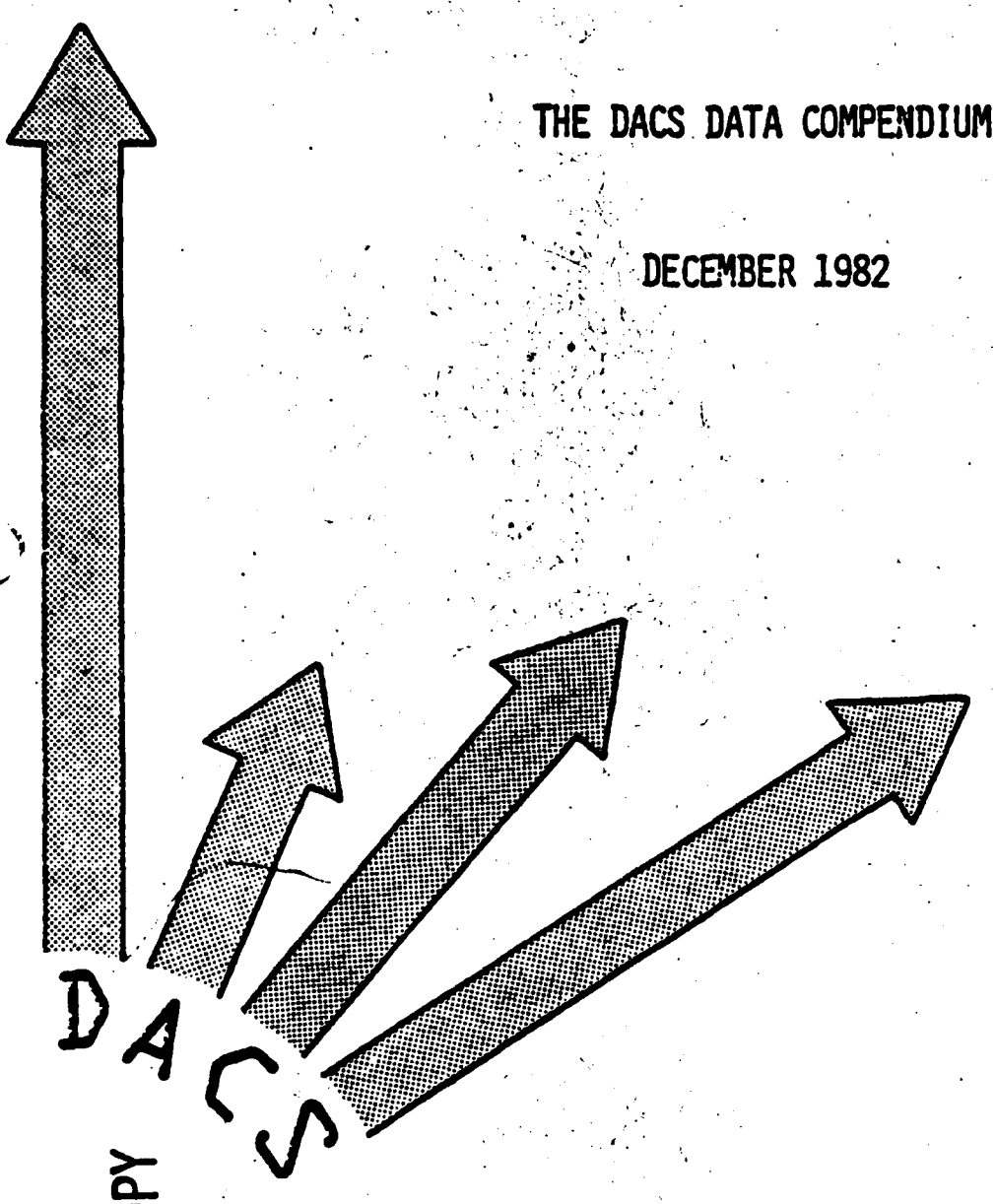
THE DACS DATA COMPENDIUM

DECEMBER 1982



Data & Analysis Center for Software

The information and data contained herein have been compiled from government and nongovernment technical reports and are intended to be used for reference purposes. Neither the United States Government nor IIT Research Institute warrant the accuracy of this information and data. The user is further cautioned that the data contained herein may not be used in lieu of other contractually cited references and specifications.

Publication of this information is not an expression of the opinion of The United States Government or of IIT Research Institute as to the quality or durability of any product mentioned herein and any use for advertising or promotional purposes of this information in conjunction with the name of The United States Government or IIT Research Institute without written permission is expressly prohibited.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: DDCP	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE *(and Subtitle)*: THE DACS DATA COMPENDIUM	5. TYPE OF REPORT & PERIOD COVERED: Interim Report April 1982 - Dec. 1982
	6. PERFORMING ORG. REPORT NUMBER: N/A
7. AUTHOR(s): Christopher S. Turner	8. CONTRACT OR GRANT NUMBER(s): F30602-81-C-0280
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Data & Analysis Center for Software RADC/ISISI Griffiss AFB, NY 13441	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Rome Air Development Center (COEE) Griffiss AFB, NY 13441	12. REPORT DATE: December 1982 13. NUMBER OF PAGES: 84
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*: Same	15. SECURITY CLASS. *(of this report)*: UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: John Palaimo (COEE)
DACS Source Code No. 413570
Cost: $10.00

Available from:
Data & Analysis Center for Software
RADC/ISISI
Griffiss AFB, NY 13441

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Software Engineering
Software Experience Data
Computer Software
Data Collection
Data Repository

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

One of the primary functions of the DACS is the acquisition, maintenance, and dissemination of empirical software life cycle data for research purposes. The DACS has acquired seven sets of data from various sources and maintains this data in the Software Life Cycle Empirical Database (SLED). In order to provide researchers with information detailed enough to determine if these datasets, a subset of one of them, or a combination of them will serve their research needs, the DACS has published this compendium of the data contained in the SLED. The Compendium provides for each dataset the time period represented by the data,

(over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

the life cycle phases represented, the parameters present, the degree of completeness of the dataset, the type (when available) of software against which the data was collected, and the number of records of each type. Seven appendices provide detailed record formats for each dataset. The Compendium also contains complete details on the procedures required for users to obtain subsets of this data in hard copy or magnetic tape format.

DACS Data & Analysis Center for Software
AN INFORMATION ANALYSIS CENTER

THE DACS DATA COMPENDIUM

DECEMBER 1982

CHRISTOPHER S. TURNER
IIT RESEARCH INSTITUTE

UNDER CONTRACT TO:

ROME AIR DEVELOPMENT CENTER
GRIFFISS AFB, NY 13441

The Data & Analysis Center for Software (DACS) is a Department of Defense (DoD) information analysis center sponsored by the Air Force Systems Command, Rome Air Development Center (RADC), and operated by IIT Research Institute (IITRI). DACS serves as a central source for current, readily usable data and information concerning software technology.

The major functions of the DACS are: to maintain a computer database of empirical data collected on the development and maintenance of computer software; to produce and distribute subsets of the database for use by software researchers; to maintain a software technology information base of technical documents, project status information, and evaluation data pertinent to the computer software field; to analyze the data and information and produce technical reports; to maintain a current awareness program which includes dissemination of technical information, assessments of technological developments, and publication of a quarterly newsletter; to provide technical assistance in the form of technical information and special studies of topics related to software engineering and software technology.

To obtain more information on the products and services of the DACS, contact:

Data & Analysis Center for Software
RADC/ISISI
Griffiss Air Force Base, NY 13441

Telephone: 315/336-0937
Autovon: 587-3395

In Dedication

This data compendium and the design of the Software Life-cycle Empirical Database (SLED) which it describes was the result of much research and hard work by Christopher S. Turner, a Programmer Analyst at the DACS. In January of 1983 Chris was taken from us as the result of a tragic automobile accident. His loss has saddened his fellow workers, and he will be missed for a long time to come.

The amount of effort he put forth in his unselfish contributions toward designing the SLED and completing the compendium has resulted in a valuable research tool being made available to the software community. We are pleased to publish this compendium in dedication to Chris's memory.

TABLE OF CONTENTS

LIST OF FIGURES

SECTION I

INTRODUCTION

The Data & Analysis Center for Software (DACS) is a DoD Information Analysis Center (IAC) which was established in response to a well recognized need for a facility to serve as a centralized source for current, readily usable data and information concerning software technology. One of the primary functions of the DACS is the acquisition, maintenance, and dissemination of empirical software life cycle data for research purposes. To this end, this document describes the data which has been acquired and which is being maintained at the DACS and delineates the procedures required for users to obtain subsets of this data.

BACKGROUND

The DACS has acquired seven sets of data from various sources and maintains this data in the Software Life Cycle Empirical Database (SLED). In that each set of data was the result of a data collection effort which pursued individually specific objectives, the resulting datasets differ with regard to:

- The time period represented by projects in a dataset
- The portion of the software life cycle represented by the data
- The aspects of the software development and/or maintenance processes measured by the data collection activity
- The quality of the data as reflected in the verification and validation procedures used in data collection
- The subsequent analyses supported by the data.

For reasons of clarity, this discussion of the contents of the SLED will be organized by contributing organizations. The titles of some of the datasets reflect their contents as opposed to the contributing organization. The seven sets of data and the sources associated with them are provided below.

(1) The DACS Productivity Dataset - Data collected from various government and private industry sources and compiled by Richard Nelson of RADC.

(2) The Reliability Dataset - Data collected at Bell Laboratories, Whippany, N.J. and compiled by John Musa.

(3) The NASA/SEL Life Cycle Dataset - Data collected and contributed by the Software Engineering Laboratory (SEL) at NASA Goddard Space Flight Center.

(4) The Verification & Validation (V&V) Dataset - Data collected under several Independent V&V contracts then summarized and delivered to the DACS by Logicon Incorporated.

(5) The ARF Error Dataset - Data collected and analyzed on the development of the Architecture Research Facility (ARF) at the Naval Research Laboratories (NRL) by David Weiss.

(6) The Baseline Software Dataset (BSDS) - Data collected on six defense software projects from various organizations.

(7) The Operations and Maintenance O&M Dataset - Data collected on the operations and maintenance of the PAVE PAWS radar system.

The datasets were generated at different points in time and it is important to consider this when analyzing data of this nature. Figure 1-1 illustrates the periods of time represented by data in each of the datasets. Each of the datasets contains data from various software life cycle phases as depicted in Figure 1-2. Of these seven datasets, four are available in a standard format. The remaining three, the NASA/SEL, the BSDS and the O&M datasets, because of their extensive nature, have not been processed into a form which is readily usable and as such distribution is limited to customized versions of these datasets, prepared in response to individual requests. Each of the seven sets of data is discussed

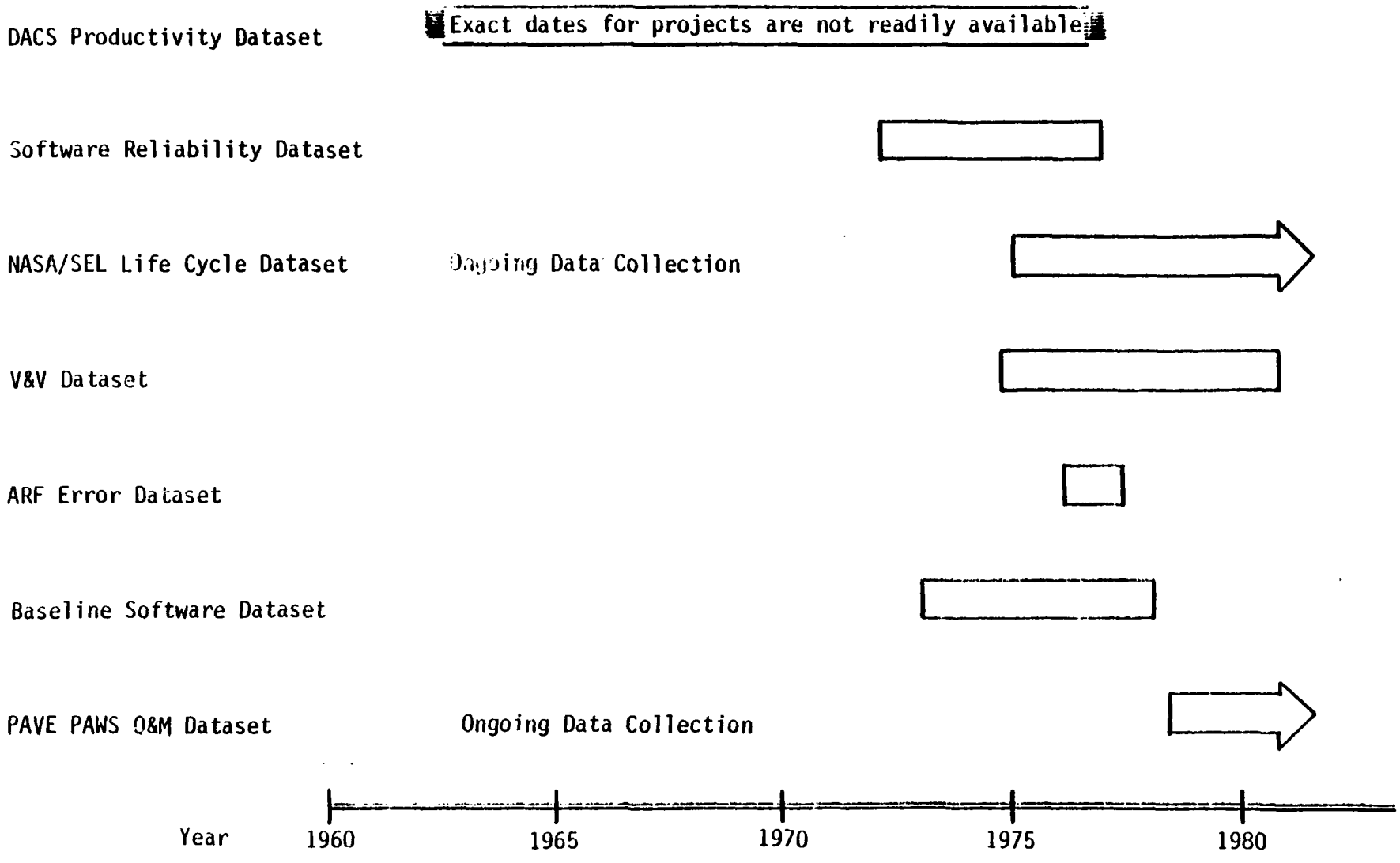


FIGURE 1-1 TIME PERIODS REPRESENTED BY SLED DATASETS

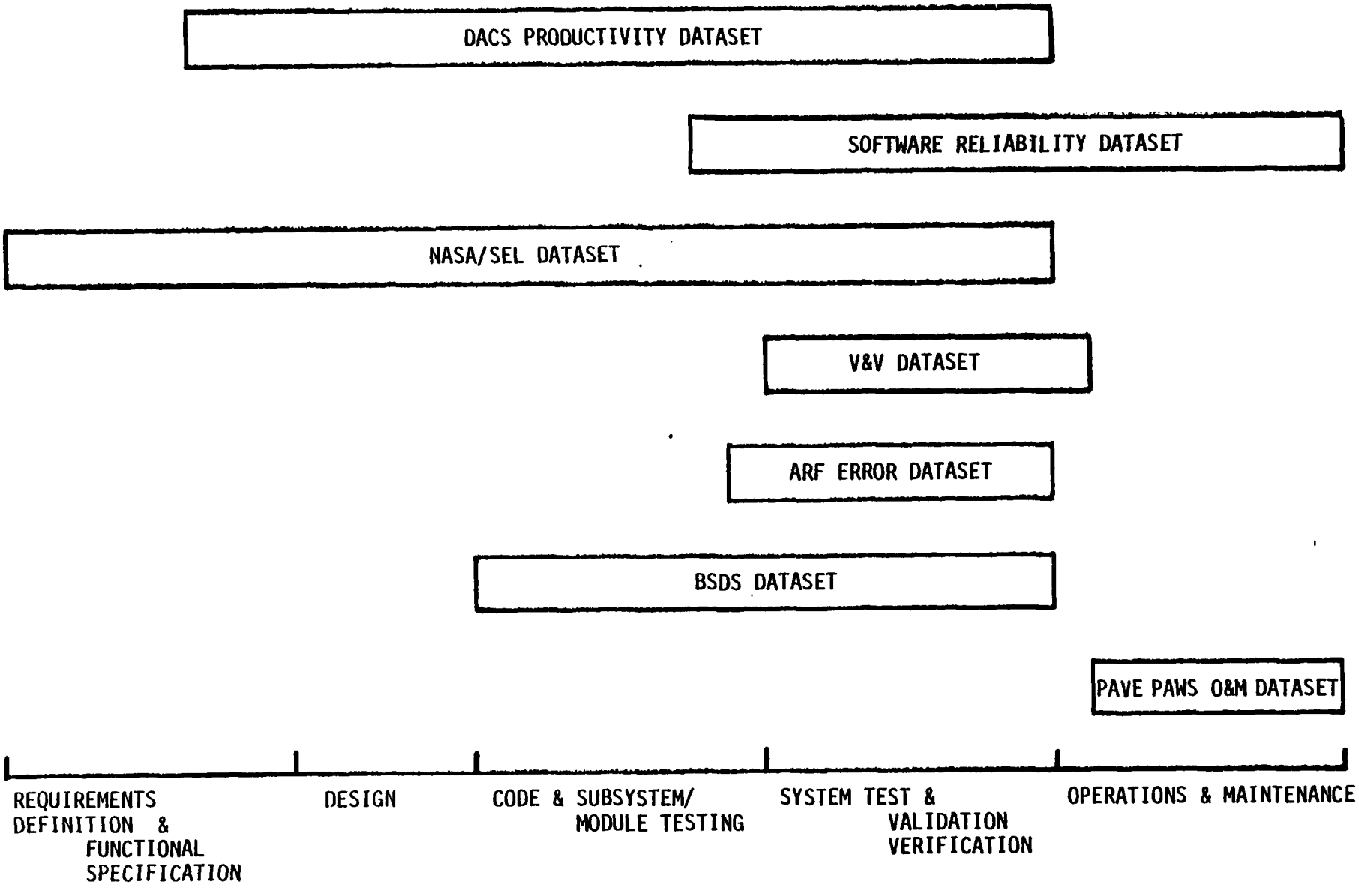


FIGURE 1-2 LIFE CYCLE PHASES DESCRIBED BY DACS DATASETS

individually in the following sections. Procedures for obtaining data from the SLED are delineated in Section 3.

SECTION II

THE DACS PRODUCTIVITY DATASET

This dataset consists of summary data on roughly 400 software projects and was compiled by Richard Nelson of RADC, [NELSON-78] whose intent was to collect data supporting analysis of the effect of implementation methodologies on productivity during software development. The data was obtained from open literature and private sources in industry and government and represents software development projects dating from the early 1960's through the mid 1970's. The software applications range from avionics and space-flight command and control functions and radar system support, to off-the-shelf software packages, communications software, and management information systems. Most of the projects represent DoD or other government applications.

Essentially eight parameters are used to describe each software project in the dataset. These parameters and their associated definitions* are provided below:

(1) Project Identification - An internally assigned identification number.

(2) Project Size - Number of lines of source code (DSLOC) in the delivered product. Source lines are 80 character source records provided as input to a language processor. Job control language, link edit language, data declarations, internal program data and comment lines are included in the count. Unmodified reused code, test drivers, "throwaway" code, and external data are not included. Where the number of lines of source code are known, but

*The definitions for these parameters are reproduced from "Software Data Collection and Analysis" Draft Report by Richard Nelson, RADC, 1978 [NELSON-78].

the content of the source code is not, the assumption has been made that the code meets the above definitions. Where the size of the code has been given in computer words, an arbitrary conversion has been made to DSLOC by:

DSLOC = size (computer words)/2 (for high order language)

DSLOC = size (computer words) (for assembly language)

(3) Project Effort - Total effort in man-months (TMM) required to produce the software product. Includes effort in management, administration, analysis, operational support, and other areas such as documentation, program design, coding and test, i.e. - all effort chargeable to the project by the builder.

(4) Project Duration - Duration of project in total months (TM) derived from begin and end dates of project, less any "dead time" in the project, e.g., work stoppage.

(5) Source Code Languages - Programming languages used on the project are recorded by name, and are expressed as a percentage of the total DSLOC written in each different language, e.g., COBOL 80%, ASSEMBLY 15%, JCL 5%. Where a program design language (PDL) has been used to define the program, the total number of PDL lines written is included as a separate quantity and is not included in the DSLOC count.

(6) Errors - Errors (ERRS) are totalled by counting the number of formally recorded Software Problem Reports (SPR) for which a fix has been generated during the period covered by the project. Redundant SPR's are assumed to have been eliminated, i.e., multiple SPR's reporting the same problem. (In general, equating errors to SPR's on a one-to-one basis produces a low count for all errors encountered during the software development process since SPR's are usually not generated until such time as the software comes under formal configuration control. This point is usually marked by the end of software unit testing and the beginning of integration testing, hence many of the errors detected during coding and subsequent compilation have already been removed. Collecting data on errors on the coding portion of the development process is impractical however, and error data derived from SPR data can be extrapolated to the entire development process.)

(7) Documentation - Delivered pages of documentation (DOC) includes program listings, flow charts (low and hi-level), operating procedures, maintenance procedures and any other descriptive material covering the design, development, test, operation, installation and maintenance of the software.

(8) Implementation - The implementation techniques used on the software project are recorded and expressed as a percentage of the DSLOC built using the specific techniques of Structured Coding, (SC), Top-Down Design and Programming (TD), Chief

Programmer Team (CPT), Code Reviews or Inspections (CR), and Librarian or Program Support Library (LIB). For definitions of the above, refer to RADC TR 74-300, Vols 1-16, "Structured Programming Series."

Additionally, several other data-items, derived from those mentioned previously, are recorded for projects in the dataset. These are:

Productivity = DSLOC/TMM

Average Number of Personnel = TMM/TM

Error Rate = ERRS/DSLOC

Error Rate (temporal) = ERRS/TMM

Documentation Rate = DOC/DSLOC

Not every parameter has been recorded for every project in the dataset. Figure 2-1 displays the number of project records where each of six primary data-items are recorded. This figure also displays the number of projects represented by combinations of two of these parameters. The record formats for this dataset are included in Appendix A.

Parameter	Number of Projects where Parameter is Recorded
Size (DSLOC)	403
Effort (TMM)	381
Languages (LANG)	374
Schedule (TM)	308
Documentation (DOC)	253
Errors (ERRS)	30

Total Number of Projects in Dataset:

Note: Numbers recorded in the boxes represent the number of projects in the dataset where both the row and column parameters, specifying a particular box are recorded. (e.g. 379 projects contain both TMM and DSLOC information)

	DSLOC	TMM	LANG	TM	DOC
	379				
	370	354			
	307	303	308		
	252	247	253	251	
	30	17	9	17	4

FIGURE 2-1 A CHART DISPLAYING THE NUMBER OF PROJECTS IN THE DACS PRODUCTIVITY DATASET CONTAINING SPECIFIC PARAMETERS AND COMBINATIONS OF PARAMETERS

Modified from a chart appearing in [NELSON-78]

THE SOFTWARE RELIABILITY DATASET

This dataset consists of software failure data on 16 projects and was compiled by John Musa of Bell Telephone Laboratories whose objective was to collect failure interval data to assist software managers in monitoring status and predicting schedules. Careful controls were employed during data collection to ensure that the data would be of high quality. The data was collected throughout the mid 1970's and represents projects of a variety of applications including real time command and control, word processing, commercial and military.

For each software failure in the dataset the following items are recorded:

(1) Project Identification - An internally assigned identification number.

(2) Failure Number - A number identifying a particular failure. Failures are consecutively numbered from the first failure reported.

(3) Failure Interval - The time elapase from the previous failure to the current failure in seconds. For one of the projects in the dataset this time is given in CPU seconds, for the remaining projects this time is given in wall-clock seconds.

(4) Day of Failure - Represents the day on which the failure occurred in terms of the number of working days from the start of the current phase or data collection period.

Figure 2-2 displays the size of each sample of failure data for each project in the dataset, as well as other useful information concerning the individual projects. More detailed information on the specific characteristics of each project is available in a report entitled "Software Reliability Data" by John Musa, Bell Telephone Laboratories [MUSA-79]. The record formats for this dataset are provided in Appendix B.

System Code	Nature of System	Size (Delivered Object Code Instructions)	Size of Failure Sample (Number of Records)	Phases Represented by Sample
1	Real Time Command & Control	21,700	136	System Test & Operations
2	Real Time Command & Control	27,700	54	System Test & Operations
3	Real Time Command & Control	23,400	38	System Test & Operations
4	Real Time Command & Control	33,500	53	System Test & Operations
5	Real Time Commercial	2,445,000	831	System Test *
6	Commercial Subsystem	5,700	73	Subsystem Test
14C	Real Time	(Hundreds of Thousands)	36	Operations *
17	Military	61,900	38	System Test
27	Military	126,100	41	System Test
40	Military	180,000	101	System Test
SS1A	Operating System	(Hundreds of Thousands)	112	Operations *
SS1B	Operating System	(Hundreds of Thousands)	375	Operations *
SS1C	Operating System	(Hundreds of Thousands)	277	Operations *
SS2	Time Sharing System	(Hundreds of Thousands)	192	Operations *
SS3	Word Processing System	(Hundreds of Thousands)	278	Operations *
SS4	Operating System	(Hundreds of Thousands)	196	Operations *

NOTE: Starred phases indicate the failure sample is not complete for that phase.

FIGURE 2-2 THE SIZE OF FAILURE SAMPLES IN THE SOFTWARE RELIABILITY DATASET

THE NASA/SEL SOFTWARE LIFE CYCLE DATASET

The Software Engineering Laboratory (SEL) was established by NASA's Goddard Space Flight Center (GSFC) in 1977 to investigate the effectiveness of software engineering techniques as applied to the development of ground-support space flight dynamics systems. The overall goals [NASA/SEL-82A] of the program are to:

(1) Understand the software development process in the GSFC environment.

(2) Measure the effects of various development models, tools, and methodologies on the development process.

(3) Identify and apply improved methodologies in the GSFC environment.

To accomplish these goals, the SEL collects extensive data on software developed by the Systems Development Section at NASA/GSFC, which is responsible for generating flight dynamics support software for GSFC-supported missions. This data is forwarded to the DACS on a roughly annual basis.

The projects represented by the data span the functions of attitude determination, attitude control, maneuver planning, orbit adjustment and general mission analysis support systems. The data collected encompasses software development projects that started as early as 1976 and projects currently under development. Figure 2-3 illustrates the history of projects covered in the dataset.

The data contained in the NASA/SEL Software Life Cycle Dataset is gathered through five sources:

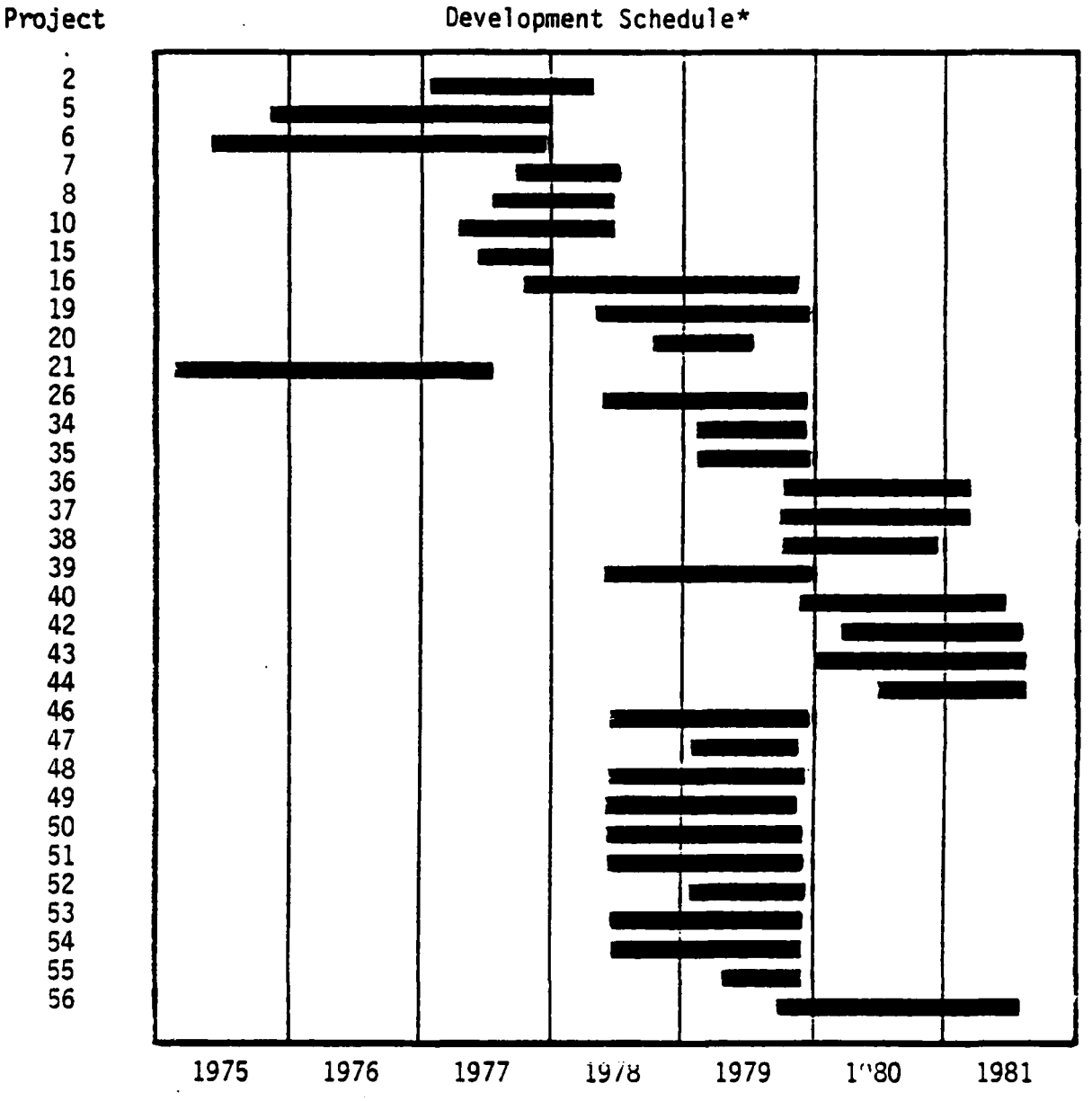


FIGURE 2-3 DEVELOPMENT TIME PERIODS FOR PROJECTS IN THE NASA/SEL DATASET

*Note: Complete scheduling date for some early projects was not recorded.

- Software Engineering Forms
- Computer Accounting
- Personal Interviews
- Automated Tools
- Management Summaries

The files in which some of the data is stored are structured according to the forms used to collect the data. These forms are summarized in Figure 2-4, reproduced from [NASA/SEL-81A]. The data collected through these forms and other media is maintained in 11 files, distributed by the DACS. These files and their contents are described below:

Encoding Dictionary File - One record is provided for each code used in the other files to represent lengthy alpha-numeric or text information. The information provided in this file, is included in record format descriptions of the other files.

Estimated Statistics File - One record is provided for each project. Summarizes size, effort, and some computer environment data. Data is actual, not estimated as implied by file name.

Header File - One record is provided for each project. Provides schedule dates for the requirements definition, design, code, system test, acceptance test and cleanup phases. Maintenance start and end dates are also provided.

Change Report File - One record is provided for each error or change occurring in a project. A record summarizes the information recorded on a Change Report form; including dates, effort required for change, type of change, and information classifying the error if the change was a result of an error.

Component Status Report File - One record is provided for each line on each Component Status Report form (completed weekly during development). Several records provide the hours spent on design, code, and test activities for each component of a given project on a weekly basis.

Component Summary File - Two records are provided for each component in each project. Records summarize a general description of each component including complexity, application, size, schedule, effort to develop, and language. Each record is either an estimate of these parameters provided when the component is first defined or the actual values of the parameters upon completion of

Form	Description of Content
General Project Summary	Computer resources used, starting and ending dates of each phase, cost information, size of product, methodologies and tools used in each phase of development, personnel involved, standards used, documentation produced, problems anticipated, and quality assurance information
Change Report	Change description, components changed, effort to change, type of change or error, and activities used to validate changes, to detect errors, and to find their cause
Resource Summary	Number of hours of worktime per week per staff member spent on a particular project, computer usage, and other charges
Component Status	Time spent during the week in a certain activity of component development (e.g., design, testing, or documentation)
Component Summary	Interfaces, programming language, complexity, resources required for each phase of development, relation to other components, and code specifications
Run Analysis	Computer used, purpose of the run, type of run, run results, and comments
Maintenance Report	Subset of change report with some maintenance-specific questions

FIGURE 2-4 NASA/SEL DATA COLLECTION FORMS

From [NASA-82A]

the project.

Resource Summary File - One record is provided for each line on the Resource Summary form (each line on the form being completed weekly by project management, up to 11 weeks per form). Records summarize the consumption of manpower, computer, support services or other resources for a specified time period.

Run Analysis File - One record is provided for each line on the Run Analysis form (up to nine computer runs per form). Records summarize the objectives and results of each computer job submitted, and whether the run was interactive or batch. Date of run is also provided.

Component Information File - One record is provided for each component in each project. Records provide information obtained from Source Analyzer Programs, including a number of Halstead complexity metrics, and instruction mix parameters.

Growth History File - One record is provided for each week during the life cycle of several NASA projects. Records provide the cumulative number of source lines written, modules and changes for each week, and are generated through an automated collection tool.

More comprehensive lists of the parameters contained in each file are provided in Appendix C. More detailed information regarding the organization and contents of the NASA/SEL Dataset may be found in [NASA/SEL-81B] and [NASA/SEL-81C].

Not every project recorded in the database contains complete data. Figure 2-5 displays the number of records in each file recorded for each project.

Project Code	NASA-SEL Files									
	EST	HF	CIF	CSF	CSR	CRF	RAF	RSF	HIS	CMT
1	1	1	191	121	392	0	216	0	0	146
2	1	1	415	225	1522	290	1164	92	42	618
3	1	1	49	0	138	0	0	0	0	0
5	1	1	683	126	1027	311	2018	99	36	1064
6	1	1	736	175	1788	491	1877	121	53	1119
7	1	1	53	55	153	55	186	11	0	217
8	1	1	539	316	810	240	984	60	25	823
9	1	1	48	22	422	0	403	0	0	165
10	1	1	944	295	1025	46	1312	91	34	423
13	1	1	53	0	79	0	0	0	0	0
15	1	1	86	0	155	0	58	0	0	36
16	1	1	12	1	46	1	0	0	0	0
19	1	1	815	865	2175	686	3168	162	53	3073
20	1	1	28	0	224	0	74	20	0	0
21	1	1	465	0	0	182	45	253	0	393
26	1	1	902	564	2407	585	2330	147	58	1574
34	1	1	46	0	392	0	2	20	0	0
35	1	1	111	0	425	103	77	63	0	213
36	1	1	530	82	5191	930	4860	211	63	5218
37	1	1	518	108	5160	751	7101	216	62	5671
38	1	1	138	179	722	0	362	93	54	290
39	1	1	83	71	502	15	111	110	0	128
40	1	1	219	67	1331	132	1017	145	52	1380
41	1	1	233	161	682	85	1	22	0	326
42	1	1	278	3	321	0	78	79	0	34
43	1	1	140	0	601	5	793	31	0	422
44	1	1	67	0	0	0	0	15	0	0
45	1	1	477	1	2011	5	0	165	46	14
57	1	1	129	1	2657	67	0	124	32	0
58	1	1	211	0	538	125	0	43	0	91
59	1	1	81	0	183	8	0	36	44	25
60	1	1	65	0	442	24	0	30	43	58
61	1	1	826	0	2145	41	5	113	44	3
62	1	1	349	0	649	6	0	88	39	23
63	1	1	1	0	113	0	0	12	0	0

File Codes:

- EST - Estimated Statistics File
- HF - Header File (Scheduling Data)
- CIF - Component Information File
- CSF - Component Summary File
- CSR - Component Status Report File
- CRF - Change Report File
- RAF - Run Analysis File
- RSF - Resource Summary File
- HIS - Growth History File
- CMT - Comment File

FIGURE 2-5 CONTENTS OF THE NASA/SEL DATASET BY PROJECT

THE VERIFICATION AND VALIDATION (V&V) DATASET

This dataset contains data collected during the independent Verification and Validation (V&V) of five software projects. The purpose of this data collection effort was to record the types of errors which can be detected during independent V&V activities. Data on three of these projects has been recorded at the subsystem level as these subsystems underwent separate V&V.

The data in this dataset is maintained in two files described below:

- Summary Data File - This file contains summary level information describing each project or subsystem compiled from various sources. Information includes size, schedule and development methodology information.
- Anomaly Report Data - This file contains data describing each anomaly or error uncovered by the independent V&V team. Anomalies included occurrences of non-conformity to pre-established standards. Each anomaly report contains information specifying the anomoly location, category, and severity as well as tools used to detect and correct the anomaly.

The actual data parameters recorded in each of the two files are provided in the record formats illustrated in Appendix D.

The data contained in this dataset is complete and of very good quality due to stringent control placed on the data collection effort. Figure 2-6 summarizes the data recorded for each project or subsystem. Further background information may be obtained from [RADA-81].

System/Subsystem Code	System/Subsystem Size	Schedule Start	Schedule End	Number of Anomaly Reports	
1 A	24,000	9/74	4/76	37	
B	24,000	2/73	11/75	28	
C	41,000	10/73	1/76	142	
					TOTAL 207
2 A	24,000	1/76	6/78	48	
B	24,000	7/75	6/78	46	
C	41,000	3/76	3/78	141	
					TOTAL 235
3 A	14,000	1/75	9/76	54	
B	39,000	1/75	7/78	117	
					TOTAL 171
4	41,000	9/77	12/79	84	
5	52,000	1/74	1/81	235	

FIGURE 2-6 CONTENTS OF THE V&V DATASET

THE ARCHITECTURE RESEARCH FACILITY (ARF) ERROR DATASET

The Architecture Research Facility (ARF) was developed at the Naval Research Laboratories (NRL) to aid the rapid simulation of different computer architectures for research and evaluation purposes. Data was collected during the development of this project to evaluate development methodologies used on the project. This data was compiled by David Weiss of the NRL. Dr. Weiss was not involved in the development of the ARF but functioned as an independent collector and validator of the data. This data was provided to the DACS in hardcopy form in March 1982, to be processed and entered online.

The dataset consists of 117 error reports dealing with 143 errors isolated and corrected during ARF development. Also included in this dataset are 253 records describing each routine of which the project is comprised and several project descriptive records. The contents of each record type are described below.

- Error Report Data - one record is provided for each error report. Contains the dates the error was observed and corrected, the type of error, the routine in which the error was isolated, the effort required to isolate and correct the error, and the activities and tools used to isolate the error and validate the routine.
- Module Descriptive Data - one record is available for each routine in the system. Contains the site of the routine, number of pre-processor statements, number of comments, subjective complexity of the routine, and the function of the module or routine.
- System Descriptive Data - several records are provided for the project. Contain or describe the size and other attributes of the project, a description of the development environment, and a description of the development methodologies used on the project.

More detailed listings of the data-items contained in the ARF dataset are provided in Appendix E. The data is of generally good quality and there

are very few missing data-items. Figure 2-7 illustrates the percentage of major parameter occurrences in the dataset. Information not available in System Descriptive Data may be obtained from two sources [WEISS-78] and [ELOV-79].

Parameter	Number of Records Containing Parameter	Percent
Type of Change	117	100%
Effort of Change	116	>99%
Date Change was Determined	116	>99%
Type of Error	116	>99%
Time to Isolate Error	115	>99%
Activities Used to Isolate Error	114	>99%
Programmer Code	116	>99%
Total Number of Records	117	

FIGURE 2-7 THE OCCURRENCE OF IMPORTANT PARAMETERS IN THE ARF DATASET

THE OPERATIONS & MAINTENANCE (O&M) DATASET

The PAVE Phased Array Warning Systems (PAWS) is an over-the-horizon radar system in operation at Otis Air Force Base, Massachusetts and Beale Air Force Base, California. The software for this system is essentially the first large DoD software project developed and maintained in a comprehensive environment of software engineering tools and techniques. Parallel to the development of this system, an extensive data collection process was initiated, employing the facilities of the Program Support Library (PSL). The PSL is a portion of the PAVE PAWS software developed to manage the configuration of the remainder of the project. In addition to this the PSL collects and maintains size and change data on a continual basis. As such, a vast amount of very detailed data tracking maintenance activities is continually being collected through the operational life of the project. Full details of the data collection effort and subsequent analysis are available in [IITR-82].

The data collected against the PAVE PAWS project is maintained at the DACS in three primary forms:

- Online Data Files - consists of seven online datafiles, summarizing information collected from other sources. These are discussed individually below.
- Offline Data Files - consists of machine readable data files and reports produced by the PSL. These are stored on magnetic tape and updated monthly.
- Hardcopy Data - consists of analyses of unresolved discrepancies, minutes from configuration review board meetings, and project specification and source code data.

The most accessible and easily processed data is included in the online data files. These seven files and their respective contents are

provided below:

(1) Maintenance Activity File - This file records maintenance activities performed on the project. Contains for each approved change, the type of maintenance activity being performed, the precision of specifications for that activity, the complexity of the maintenance effort, and data concerning how the error was detected and diagnosed.

(2) CPCG Description File - Contains data providing information concerning the characteristics of the PAVE PAWS software at the CPCG level, including size of the CPCG in source lines and words of machine code, number of CPCI's, environmental factors and development constraints. (CPCG is an acronym for Computer Program Configuration Group, CPCI is an acronym for Computer Program Configuration Item.)

(3) CPCG Status File - Contains information concerning the status of CPCG's during the maintenance cycle. The size of the CPCG, its version identification, and the date the last change was made to the CPCG.

(4) Segment Change History File - Contains a history of changes or additions to the project at the segment level. The size of the segment, and the number of changes made to the segment is recorded each time a change occurred.

(5) Change History File - Contains a history of changes or additions to the project at the program level. The size of the program (in number of segments) is maintained at each change.

(6) Discrepancy Report History File - Contains information recorded on each Discrepancy Report including the origin and dates associated with actions taken.

(7) Personnel Experience Profile - Provides background information on personnel assigned to the maintenance of PAVE PAWS software, including education, work experience, programming language experience, and experience on related projects.

Detailed contents of each of these files are available in Appendix F. Figure 2-8 provides the number of records in each of these files.

The majority of data collected by the PSL is stored in essentially one file, the Configuration Management Data Base. This file consists of a number of Discrepancy Report records, describing the status of changes made to the system. Each record contains the following types of information:

File	Number of Records
Maintenance Activity File	984
CPCG Description File	64
CPCG Status File	64
Segment Change History File	7
Change History File	7
Discrepency Report History File	18
Programmer Experience Profile File	81

FIGURE 2-8 CONTENTS OF THE O&M DATASET

- A description of the discrepancy report.
- The date the discrepancy report was opened.
- The priority of action to be taken.
- Dates associated with various stages of maintenance.
- Effort expended on the maintenance activity.

More detailed discussions of the functions of the PSL are provided in [LUPINO-74] and [TINANOFF-74].

THE BASELINE SOFTWARE DATA SYSTEM (BSDS) DATASET

This DACS dataset consists of data describing six large software development efforts. These projects and the data are described in [RYE-77], [BAKER-77], [WILLIAM-77], [FRIES-77], [THAY-70], and [HECT-77]. The system application areas encompass command and control, real-time control for land-based radar, onboard guidance and navigation, and database management.

The majority of data contained in this dataset is derived from Software Problem Reports collected against the six projects. These problem reports are described in terms of:

- The dates that the problem report was opened and closed.
- The module manifesting the problem.
- The module changed to correct the problem.
- The problem category.
- The problem priority.
- The corrective action.

Three of the projects also contain data describing modules. This data includes the size of the module in source instructions, the source programming language used, the type of construction for the module, and a designation of the functional application of the module. Additionally, one project contains data describing each test run during development. Figure 2-9 summarizes the number of records and types of records provided for each project in the dataset. Detailed record formats are not available for this dataset, however, Appendix G provides a comprehensive list of parameters recorded.

Project Number	Number of Records	Number of Data Items	Number of Record Types
1	4,970	28	1
2	2,113	46	5
3	2,274	35	2
4	11,730	17	1
5	8,106	18	2
6	2,719	15	1

FIGURE 2-9 PROJECT IN THE BSDS DATASET

Reproduced from [DUVALL-79II]

THE DACS COMPOSITE PRODUCTIVITY DATASET

The SLED is currently undergoing enhancement through a restructuring of copies of portions of its contents. This will result in a database that may be searched for projects of a specific language, application or other qualifying characteristic. As part of the investigation into the feasibility of performing this effort a dataset was constructed containing data extracted from three of the datasets discussed in the previous section. The DACS Composite Productivity Dataset contains data from the DACS Productivity Dataset, the NASA/SEL Dataset, and the V&V Dataset. Essentially, the same parameters provided in the RADC dataset are provided here, with the exception of source language composition. The DACS Composite Productivity Dataset provides only the primary source language as opposed to the three primary source languages. Additionally, a field has been added to indicate from which dataset the project originated. The data contained in this dataset will be verified and additional fields and records will be added on a continuing basis. Figure 2-10 illustrates the composition of the dataset and detailed record formats are provided in Appendix H.

Original Dataset	Number of Projects
DACS Productivity	407
NASA/SEL Life Cycle	19
V&V	10
TOTAL	436

FIGURE 2-10 COMPOSITION OF THE DACS COMPOSITE PRODUCTIVITY DATASET

SECTION III

ORDERING INFORMATION

The DACS is a Defense Logistics Agency (DLA) information analysis center (IAC). As a DoD/DLA IAC the DACS is required to institute a system of user charges with the object of recovering a significant portion of the cost of operating the DACS. Charges for DACS products are designed to recover a portion of their development cost and the cost of reproduction, handling and mailing.

Orders for DACS products must be placed directly with the DACS. Orders should clearly specify the dataset(s) and services desired. Except for blanket purchase orders, prepayment is required. Please make checks payable to: IITRI/DACS.

At this time, the DACS is distributing copies of the Software Reliability Dataset, the DACS Productivity Dataset, the V&V Dataset, the ARF Error Dataset and the DACS Composite Productivity Dataset in standard formats. The price lists at the end of this compendium give prepaid prices for these standard DACS datasets. Subsets of the Baseline Software Dataset, the NASA/SEL Dataset, and the O&M Dataset are available at costs dependent upon processing time.

Call or write for assistance in determining the availability of data to meet your research needs and to determine its cost.

REFERENCES

[BAKER-77] Baker, W.F., Software Data Collection and Analysis: A Real-Time System Project History, Rome Air Development Center, RADC-TR-77-192, Griffiss Air Force Base, NY, June 1977.

[DUV-79] Duvall, Lorraine M., Baseline Software Data System, System Description, Vol. 1 (of two), Rome Air Development Center, RADC-TR-79-185, Griffiss Air Force Base, NY, July 1979.

[DUV-7911] Duvall, Lorraine M., and Christine Curtis, Baseline Software Data System, Database Reference Manual, Vol. 11 (of two), Rome Air Development Center, RADC-TR-79-185, Griffiss Air Force Base, NY, July 1970.

[ELOV-79] Elovitz, Honey S., "An Experiment in Software Engineering: The Architecture Research Facility as a Case Study," Proceeding of the Fourth International Conference on Software Engineering, September 1979.

[FRIES-77] Fries, M.J., Software Error Data Acquisition, Rome Air Development Center, RADC-TR-77-188, Griffiss Air Force Base, NY, June 1977.

[HECT-77] Hecht, H. et al., Reliability Measurement During Software Development, NASA Langley Research Center, NASA-CR-145205, Hampton, VA, September 1977.

[IITR-82] IIT Research Institute, MPP in the Pave Paws Software Maintenance Environment, RADC-TR-82-169, June 1982.

[MUSA-79] Musa, John D., "Validity of the Execution Time Theory of Software Reliability," IEEE Transactions on Reliability August 1979.

[MUSA-80] Musa, John D., Software Reliability Data, Bell Laboratories, Whippany, NJ, January 1980.

[NASA-81A] Church, V.E., et al., Guide to Data Collection, NASA Goddard Space Flight Center, Greenbelt, MD, September 1981.

[NASA-81B] Wycoff, D.C., et al., Software Engineering Laboratory (SEL) Data Base Organization and User's Guide, NASA Goddard Space Flight Center, Greenbelt, MD, September 1981.

[NASA-81C] Card, D.N., et al., Software Engineering Laboratory (SEL) Data Base Maintenance System (DBAM) User's Guide and System Description, NASA Goddard Space Flight Center, Greenbelt, MD, September 1981.

[NASA-82A] Card, D.N., McGarry, F. E., et al., The Software Engineering Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, February 1982.

[NELSON-78] Nelson, Richard, Software Data Collection and Analysis, Rome Air Development Center, Griffiss Air Force Base, NY, September 1978.

[RYE-77] Rye, P., et al., Software Systems Development: A CSDL Project History, Rome Air Force Development Center, RADC-TR-77-213, Griffiss Air Force Base, NY, June 1977.

[THAY-76] Thayer, T.A., Software Reliability Study, Rome Air Development Center, RADC-TR-76-238, Griffiss Air Force Base, NY, August 1976.

[TURN-81B] Turner, C. and G. Caron, DACS Technical Monograph Series: A Comparison of RADC and NASA/SEL Software Development Data, Data and Analysis Center for Software, Rome Air Development Center, May 1981.

[WEISS-78] Weiss, David M., Evaluating Software Development by Error Analysis: The Data from the Architecture Research Facility, NRL Report 8268, December 1978.

[WILLIAM-77] William, H.E., Jr., et al., Software Systems Reliability: A Raytheon Project History, Rome Air Developmemt Center, RADC-TR-77-188, Griffiss Air Force Base, NY, June 1977.

May 1983

			PRICE PER COPY
( )	DS-P	DACS Productivity Dataset (Magnetic Tape)	$50
( )		Hard Copy Listing	30
( )	TM-2	Software Data Collection and Analysis (Draft-Partial Report)	10
( )	DS-R	Software Reliability Dataset (Magnetic Tape)	50
( )		Hard Copy Report, "Software Reliability Data"	10
( )	DS-V	V&V Dataset (Magnetic Tape)	50
( )		Hard Copy Listing	30
( )	DS-S	Composite Productivity Dataset (Magnetic Tape)	30
( )		Hard Copy Listing	30
( )	DS-A	Architectural Research Facility (ARF) Dataset (Magnetic Tape)	50
( )		Hard Copy Listing	30

If ordering Magnetic Tapes, please check one of the formats below:*

	TRACK	DENSITY	CHARACTER SET	LABEL
☐	7	800	BCD	GCOS
☐	7	556	BCD	None
☐	7	800	BCD	None
☐	9	800	EBCDIC	None
☐	9	1600	EBCDIC	None
☐	9	800	ASCII	ANSI
☐	9	1600	ASCII	ANSI

If ordering a hard copy listing of the DACS Productivity Dataset, please specify one of the following:

Unsorted (dataset order)

( ) Sorted on programming language

( ) Sorted on delivered source code lines

( ) Sorted on project person-months

*Prices quoted for magnetic tapes include tape. If you wish to supply your own tape, deduct $15 from the quoted price and mail a blank tape with your check.

* *

ORDER FORM

Enclosed find $__________

Send order and check to:

Data & Analysis Center for Software
RADC/ISISI
Griffiss AFB, NY 13441

Phone: 315/336-0937
Autovon: 587-3395

Please send me the products/services checked above

Date ______________________________

Name/Title ______________________________

Organization ______________________________

Address ______________________________

City/State ____________________ Zip ________

Telephone ______________________________

PREPAYMENT OF ORDERS IS REQUIRED.
Checks must be made payable to IITRI/DACS.

The Data & Analysis Center for Software is a DoD Information Analysis Center
Operated by IIT Research Institute

May 1983

			PRICE PER CO
( )	DS-P	DACS Productivity Dataset (Magnetic Tape)	$50
( )		Hard Copy Listing	30
( )	TM-2	Software Data Collection and Analysis (Draft-Partial Report)	10
( )	DS-R	Software Reliability Dataset (Magnetic Tape)	50
( )		Hard Copy Report, "Software Reliability Data"	10
( )	DS-V	V&V Dataset (Magnetic Tape)	50
( )		Hard Copy Listing	30
( )	DS-S	Composite Productivity Dataset (Magnetic Tape)	30
( )		Hard Copy Listing	30
( )	DS-A	Architectural Research Facility (ARF) Dataset (Magnetic Tape)	50
( )		Hard Copy Listing	30

If ordering Magnetic Tapes, please check one of the formats below:*

	TRACK	DENSITY	CHARACTER SET	LABEL
☐	7	800	BCD	GCOS
☐	7	556	BCD	None
☐	7	800	BCD	None
☐	9	800	EBCDIC	None
☐	9	1600	EBCDIC	None
☐	9	800	ASCII	ANSI
☐	9	1600	ASCII	ANSI

If ordering a hard copy listing of the DAC Productivity Dataset, please specify one o the following:

Unsorted (dataset order)

( ) Sorted on programming language

( ) Sorted on delivered source code lines

( ) Sorted on project person-months

*Prices quoted for magnetic tapes include tape. If you wish to supply your own tape, deduct $15 from the quoted price and mail a blank tape with your check.

* *

ORDER FORM

Enclosed find $__________

Send order and check to:

Data & Analysis Center for Software
RADC/ISISI
Griffiss AFB, NY 13441

Phone: 315/336-0937
Autovon: 587-3395

Please send me the products/services checked abo

Date ______________________________

Name/Title ______________________________

Organization ______________________________

Address ______________________________

City/State ____________________ Zip ______

Telephone ______________________________

PREPAYMENT OF ORDERS IS REQUIRED.
Checks must be made payable to IITRI/DACS.

The Data & Analysis Center for Software is a DoD Information Analysis Center
Operated by IIT Research Institute

APPENDIX A

RECORD FORMATS and DATA DICTIONARY
for the
DACS PRODUCTIVITY DATASET

Record Format
DACS Productivity Dataset
Record Type A

Field	Description	Format	Comments
9- 12	Project Code	I(4)	
13	Record Type	A(1)	"A"
14- 16	Source of Data	I(3)	References Nelsons Bibliography
18- 57	Filler		
59- 69	Primary Language	A(11)	
71- 73	Percent of Primary Language	I(3)	

Record Format
DACS Productivity Dataset
Record Type B

Field	Description	Format	Comments
9- 12	Project Code	I(4)	
13	Record Type	A(1)	"B"
14- 25	Secondary Language	A(12)	
27- 29	Percent of Secondary Language Usage	I(3)	
31- 42	Tertiary Language	A(12)	
44- 46	Percent of Tertiary Language Usage	I(3)	
48- 53	Source Code Written in a PDL	I(6)	
55- 59	Ratio of PDL Source to Total	F(5.3)	
61- 66	Pages of Documentation	I(6)	
68- 72	Ratio Documentation to Total Source	F(5.3)	

Record Format
DACS Productivity Dataset
Record Type C

Field	Description	Format	Comments
9- 12	Project Code	I(4)	
13	Record Type	A(1)	"C"
14- 20	Delivered Source Lines of Code	I(7)	
22- 26	Manmonths of Development Effort	I(5)	Charged to Project
28- 30	Development Schedule Length	I(3)	To nearest month
32- 36	Productivity (Lines per Manmonth)	I(5)	
38- 40	Structured Code Usage	I(3)	Percent of total source
42- 44	Top-Down Programming Usage	I(3)	Percent of total source
46- 48	Chief Programmer Team Usage	I(3)	Percent of total source
50- 52	Programmer/Librarian Usage	I(3)	Percent of total source
54- 56	Formal Code Review Usage	I(3)	Percent of total source
58- 62	Number of Software Problem Reports	I(5)	
64- 69	Ratio of SPRs to Total Source	F(6.3)	

APPENDIX B

RECORD FORMAT for the SOFTWARE RELIABILITY DATASET

Record Format
Software Reliability Dataset
Failure Interval Record

Field	Description	Format	Comments
1- 4	Project Code	A(4)	
5- 9	Failure Number	I(5)	Sequentially assigned
10- 19	Failure Interval Length	I(9)	In seconds (see note)
20- 24	Day of Failure	I(5)	From start of project phase or data collection period

Note
Failure intervals for project 6 are in CPU seconds; wall clock time is used for all other projects

APPENDIX C

RECORD FORMATS for the NASA/SEL DATASET

Record Format
NASA-SEL Software Engineering Dataset
Component Information File (CIF)

Field	Description	Format	Comments
1- 2	Project Code	I(2)	
3- 10	Filler		
11- 13	Component Code	I(3)	
14- 15	Panvalet Level Number	I(2)	
16- 17	Module Function	I(2)	From Encoding Dictionary 1 = I/O processing 2 = Algorithmic
18- 19	System/Subsystem Function	I(2)	3 = Logic control 4 = System related 5 = Data/COMMON block 6 = Other Blank = no response
20	Origin	I(1)	From Encoding Dictionary 1 = New code 2 = Extensively modified code 3 = Slightly modified code 4 = Copy of existing code
21- 24	Number of Executable Statements	I 4)	
25- 28	Number of Lines with Comments	I(4)	
29- 31	Number of Comment Lines	I(4)	
32- 34	Number of Unique Operators	I(3)	
35- 37	Number of Unique Operands	I(3)	
38- 41	Total Number of Operators	I(4)	
42- 45	Total Number of Operands	I(4)	
46- 48	Number of I/O Variables from Module	I(3)	
49- 51	Number of Decisions	I(3)	McCabes measure
52- 54	Number of FUNCTION References	I(3)	
55- 57	Number of I/O Statements	I(3)	
58- 60	Number of Assignment Statements	I(3)	
61- 63	Number of CALL Statements	I(3)	
64- 66	Number of FORMAT Statements	I(3)	
67	Status Flag	I(1)	1 = Unchecked from GSFC 2 = Unchecked from UM 5 = Hand checked

Record Format
NASA-SEL Software Engineering Dataset
Comment File (CMT)

Field	Description	Format	Comments
1- 6	Form Number	A(6)	Eg. D00633
7- 8	Sequence Number	I(2)	
9	Comment Type	A(1)	C = Comment D = Description R = Reason U = Useful item
11- 12	Project Code	I(2)	
13	Comment is Continued	A(1)	Y = Yes N = No
14-103	Text of Comment	A(90)	
104	Status Flag	I(1)	1 = Unchecked from GSFC 2 = Unchecked from UM 5 = Hand checked

Record Format
NASA-SEL Software Engineering Dataset
Change Report File (CRF)

Field	Description	Format	Comments
1- 6	Form Number	A(6)	Eg. D00633
7- 8	Project Code	I(2)	
9- 13	Programmer Code	I(5)	
14- 19	Form Date	I(6)	YYMMDD
20- 21	Number of Components Changed	I(2)	May be greater than 5
22- 23	Number of Components Examined	I(2)	
24	More than One Component Affected	I(1)	To be added
25- 30	Date Change was Determined	I(6)	YYMMDD
31- 36	Date Change was Started	I(6)	YYMMDD
37	Effort for Change	I(1)	From Encoding Dictionary 1 = Less than one hour 2 = One hour to a day 3 = One day to three days 4 = Over three days Blank = no response
38- 41	Type of Change	4A(1)	From Encoding Dictionary Up to 4 responses 1 = Error correction 2 = Planned Enhancement 3 = Implement Req. change 4 = Improve clarity 5 = Improve user service 6 = Develop utility only 7 = Optimization 8 = Adapt to envir. change 9 = Other Blank = no response
42- 56	Codes of changed Components	5I(3)	Up to five component codes
57- 60	Type of Error	4I(1)	From Encoding Dictionary Up to four responses 1 = Req. incorrect 2 = Funct. Specs. incorrect 3 = Design error several comps. 4 = Design error one comp. 5 = Misunderstand ext. envir. 6 = Error in language usage 7 = Clerical error 8 = Other Blank = no response
61	Phase when Error Entered System	I(1)	From Encoding Dictionary 1 = Requirements Def. 2 = Functional Specification 3 = Design 4 = Code and Test 5 = System Test 6 = Unknown Blank = no response
62	Data Structure Error Flag	A(1)	X = Yes Blank = No
63	Control Logic Error Flag	A(1)	X = Yes Blank = No

Record Format
NASA-SEL Software Engineering Dataset
Change Report File (CRF) continued

Field	Description	Format	Comments
	ERROR ISOLATION ACTIVITIES		From Encoding Dictionary
64- 68	For Program Validation	5A(1)	Up to five response each 1 = Pre-acceptance test 2 = Acceptance test
69- 73	For Detecting Symptoms	5A(1)	3 = Post-acceptance test 4 = Inspection of output 5 = Code reading by prgmr.
74- 78	Tried in Finding Cause	5A(1)	6 = Code reading by another 7 = Talk with other prgmrs. 8 = Special debug code
79- 83	For Finding Cause	5A(1)	9 = System error message A = Project specif. err. msg. B = Reading Documentation C = Trace D = Dump E = Cross reference F = Proof Technique G = Other Blank = no response
84	Time to Isolate Error	I(1)	From Encoding Dictionary 1 = Less than one hour 2 = One hour to a day 3 = More than one day 4 = Never found Blank = no response
85	Work Around Used Flag	A(1)	Y = Yes N = No Blank = no response
86	Related to Previous Change Flag	A(1)	Y = Yes N = No Blank = no response
87- 91	Previous Form Number	I(5)	Excludes first character, includes leading zeroes eg. 00633
92- 97	Previous Form Date	I(6)	YYMMDD
98	Reason Comment Flag	A(1)	Y = Yes N = No
99	Descriptive Comment Flag	A(1)	Y = Yes N = No
100	General Comment Flag	A(1)	Y = Yes N = No
101	Status Flag	I(1)	1 = Unchecked from GSFC 2 = Unchecked from UM 5 = Hand checked

Record Format
NASA-SEL Software Engineering Dataset
Component Summary File (CSF)

Field	Description	Format	Comments
1- 6	Form Number	A(6)	Eg. I00633
7- 8	Project Code	I(2)	
9- 13	Programmer Filling out Form	I(5)	Programmer Code
14- 18	Programmer Implementing Component	I(5)	Programmer Code
19- 24	Form Date	I(6)	YYMMDD
25	Form Stage	A(1)	N = New U = Under development C = Complete
26- 28	Component Code	I(3)	
29	Precision of Specification	I(1)	From Encoding Dictionary 1 = Very Precise 2 = Precise 3 = Imprecise Blank = no response
30	Complexity	A(1)	E = Easy M = Moderate H = Hard Blank = no response
31- 33	Type of Software	3I(1)	From Encoding Dictionary Up to 3 responses 1 = I/O processing 2 = Algorithmic 3 = Logic control 4 = Systems related 5 = Data/COMMON block 6 = Other Blank = no response
34- 36	Percent Assignment Statements	I(3)	
37- 39	Percent Control Statements	I(3)	
40- 42	Percent Other Statements	I(3)	
43- 47	Number of Stmts w/o Comments	I(5)	
48- 52	Number of Stmts w/ Comments	I(5)	
53- 57	Number of Machine Bytes	I(5)	
58	Independent of Other Software Flag	A(1)	Y = Yes N = No Blank = no response
59	Relation to Other Software	I(1)	From Encoding Dictionary 1 = Inserted at lower level 2 = New driver or interface 3 = Redesign existing comps. 4 = Rename existing comps. 5 = Regroup exist. material 6 = Other Blank = no response
60- 63	Type of Addition to Project	4A(1)	From Encoding Dictionary Up to 4 responses 1 = Error correction 2 = Planned enhancement 3 = Implement Req. change 4 = Improve clarity

Record Format
NASA-SEL Software Engineering Dataset
Component Summary File (CSF) continued

Field	Description	Format	Comments
60- 63	Type of Addition to Project continued	4A(1)	5 = Improve user service 6 = Develop utility only 7 = Optimization 8 = Adapt to envir change 9 = Other Blank = no response
64- 65	Number of Components Called	I(2)	
66	Filler		
67- 68	Number of Calling Components	I(2)	
69	Filler		
70- 71	Number of Shared Items	I(2)	
72	Filler		
73- 74	Number of Components Descending	I(2)	
75	Filler		
76- 77	Primary Language Used	I(2)	From Encoding Dictionary 1 = FORTRAN 2 = Assembly Blank = no response
78- 80	Percent of Primary Language	I(3)	
81- 82	Secondary Language Used	I(2)	From Encoding Dictionary 1 = FORTRAN 2 = Assembly Blank = no response
83- 85	Percent of Secondary Language	I(3)	
	FORM OF DESIGN		
86- 87	Levels Using Functional Design	2I(1)	Up to 2 responses each
88- 89	Levels using Procedural Design	2I(1)	1 = Component level
90- 91	Levels Using English Design	2I(1)	2 = Subcomponent level
92- 93	Levels Using Formal Design	2I(1)	3 = Basic block segment
94- 95	Levels Using Other Design Form	2I(1)	4 = Statement level 5 = Other Blank = no response
	CONSTRAINTS		
96- 97	Memory Space Constraint	A(2)	Responses for each
98- 99	Execution Time Constraint Flag	A(2)	X = Yes
100-101	Other Constraint Flag	A(2)	Blank = No
102-104	Number of Design Runs on Computer	I(3)	
105-107	Number of Code Runs on Computer	I(3)	
108-110	Number of Test Runs on Computer	I(3)	
111-113	Computer Time for Design	F(3.1)	Tenths of CPU minutes
114-116	Computer Time for Code	F(3.1)	Tenths of CPU minutes
117-119	Computer Time for Test	F(3.1)	Tenths of CPU minutes
120-122	Effort for Design	F(3.1)	Tenths of manhours
123-125	Effort for Code	F(3.1)	Tenths of minutes
126-128	Effort for Test	F(3.1)	Tenths of manhours
129-134	Estimated Design Phase End Date	I(6)	YYMMDD
135-140	Estimated Coding Phase End Date	I(6)	YYMMDD
141-146	Estimated Testing Phase End Date	I(6)	YYMMDD
147	Description Comment Flag	A(1)	Y = Yes N = No

Record Format
Nasa-SEL Software Engineering Dataset
Component Summary File (CSF) continued

Field	Description	Format	Comments
148-162	Components Called	5I(3)	Up to 5 component codes
163-177	Calling Components	5I(3)	Up to 5 component codes
178-192	Shared Components (Items)	5I(3)	Up to 5 component codes
193-207	Components affected by Reorganization	5I(3)	Up to 5 component codes
208-227	Other Form of Design Name	A(20)	
228-247	Other Constraint Name	A(20)	
248	Useful Items Comment Flag	A(1)	Y = Yes N = No
249	Additional Comment Flag	A(1)	Y = Yes N = No
250	Status Flag	A(1)	1 = Unchecked from GSFC 2 = Unchecked from UM 5 = Hand checked

Record Format
NASA-SEL Software Engineering Dataset
Component Status Report File (CSR)

Field	Description	Format	Comments
1- 6	Form Number	A(6)	Eg. B00952
7- 8	Sequence Number	I(2)	
9- 10	Project Code	I(2)	
11- 15	Programmer Code	I(5)	
16- 21	Form Date	I(6)	YYMMDD
22- 24	Component Code	I(3)	
25- 60	Hours Spent in Each Activity	9F(4.1)	Up to 9 responses in tenths of manhours Activities: 1 Design creation 2 Design reading 3 Design review 4 Code development 5 Code reading 6 Code reviewing 7 Module testing 8 Integration testing 9 Test reviewing
61- 68	Other Activity Name	A(8)	
69- 72	Hours Spent in Other Activity	F(4.1)	Tenths of manhours
73	Status Flag	I(1)	1 = Unchecked from GSFC 2 = Unchecked from UM 5 = Hand checked
74	Phase Flag	A(1)	R = Requirements D = Development M = Maintenance

Record Format
NASA-SEL Software Engineering Dataset
Encoding Dictionary File (ENC)

Field	Description	Format	Comments
1- 3	Code Type	I(3)	Numeric code identifying the category
4- 8	Code	A(5)	Alphanumeric code identifying a particular value
9- 16	Abbreviation	A(8)	Eg. JCLERROR
17- 60	Verbal Description	A(44)	

Record Format
NASA-SEL Software Engineering Dataset
Estimated Statistics File (EST)

Field	Description	Format	Comments
1- 8	Filler		
9- 10	Project Code	I(2)	
11- 14	Total Number of Components	I(4)	
15- 18	Total Number of Modules	I(4)	
19- 22	Number of New Modules	I(4)	
23- 26	Number of Modified Modules	I(4)	
27- 32	Number of Computer Runs	I(6)	
33- 38	Number of Source Code Changes	I(6)	
39- 44	Number of Pages of Documentation	I(6)	
45- 50	Total Number of Lines of Source Code	I(6)	
51- 56	Number of New Source Code Lines	I(6)	
57- 62	Number of Modified Source Lines	I(6)	
63- 68	Total Number of Executable Statements	I(6)	
69- 74	Number of New Executable Statements	I(6)	
75- 80	Number of Modified Executable Stmts	I(6)	
81- 86	Programmer Work Hours	F(6.1)	In tenths of manhours
87- 92	Management Work Hours	F(6.1)	In tenths of manhours
93- 98	Other (Services) Work Hours	F(6.1)	In tenths of manhours
99-104	IBM 360-95 Computer Hours	F(6.1)	In tenths of hours
105-110	IBM 360-75 Computer Hours	F(6.1)	In tenths of hours
111-116	Other Computer Hours	F(6.1)	In tenths of hours
117	Status Flag	I(1)	1 = Unchecked 2 = Hand checked 3 = Verified by application
118	Active Flag	A(1)	Y = Active N = Inactive Blank = no response
119	Project Category	I(1)	1 = Attitude oriented 2 = Orbit oriented 3 = Scientific oriented 4 = Database oriented 5 = Tool 6 = Real time 7 = Other Blank = No response

Record Format
NASA-SEL Software Engineering Dataset
Phase Date File (HDR)

Field	Description	Format	Comments.
1- 8	Filler		
9- 10	Project Code	I(2)	
11- 12	Development Computer	I(2)	1 = IBM 360 2 = PDP 11/70 Blank = no response
13- 14	Target Computer	I(2)	1 = IBM 360 2 = PDP 11/70 Blank = no response
15	Extent of Alien Computer Use	I(1)	To be added
	PHASE DATES		
16- 21	Requirements Definition Start	I(6)	YYMMDD
22- 27	Requirements Definition End	I(6)	YYMMDD
28- 33	Design Start	I(6)	YYMMDD
34- 39	Design End	I(6)	YYMMDD
40- 45	Code and Test Start	I(6)	YYMMDD
46- 51	Code and Test End	I(6)	YYMMDD
52- 57	System Test Start	I(6)	YYMMDD
58- 63	System Test End	I(6)	YYMMDD
64- 69	Acceptance Test Start	I(6)	YYMMDD
70- 75	Acceptance Test End	I(6)	YYMMDD
76- 81	Cleanup Start	I(6)	YYMMDD
82- 87	Cleanup End	I(6)	YYMMDD
88- 93	Maintenance Start	I(6)	YYMMDD
94- 99	Maintenance End	I(6)	YYMMDD
100-111	Spares	A(12)	
112	Status Flag	I(1)	1 = Unchecked from GSFC 2 = Unchecked from UM 5 = Hand Checked

Record Format
NASA-SEL Software Engineering Dataset
Growth History File (HIS)

Field	Description	Format	Comments
1- 2	Project Code	I(2)	
3- 8	Date	I(6)	YYMMDD
9- 14	Number Commented Source Lines to Date	I(6)	
15- 17	Number of Modules to Date	I(3)	
18- 23	Number of Changes to Date	I(6)	

Record Format
NASA-SEL Software Engineering Dataset
Run Analysis File (RAF)

Field	Description	Format	Comments
1- 6	Form Number	A(6)	Eg. J00633
7- 8	Sequence Number	I(2)	
9- 10	Project Code	I(2)	
11- 15	Programmer Code	I(5)	
16- 21	Run Date	I(6)	YYMMDD
22- 23	Computer Code	I(2)	To be added
24	Interactive Flag	A(1)	X = Interactive Blank = no response
25- 28	Run Purpose	I(1)	From Encoding Dictionary 1 = Unit test 2 = System test 3 = Benchmark test 4 = Maintenance or utility 5 = Compile, Assm., or link 6 = Debug run 7 = Other Blank = no response
29- 30	Number of Components	I(2)	May be more than 5
31- 45	Component Codes	5I(3)	Up to 5 component codes
46	First Run Flag	A(1)	X = First run Blank = no response
47	Run Met Objectives	A(1)	Y = Yes N = No Blank = no response
48- 51	Run Results	4A(1)	From Encoding Dictionary Up to 4 responses 1 = Good run 2 = Submit error 3 = JCL error 4 = Other setup error 5 = Hardware error 6 = Software error 7 = Compile error 8 = Link error 9 = Execution error A = User generated message B = Ran to completion Blank = no response
52	Comment Indicator Flag	A(1)	Y = Yes N = No
53	Status Flag	I(1)	1 = Unchecked from GSFC 2 = Unchecked from UM 5 = Hand checked

Record Format
NASA-SEL Software Engineering Dataset
Resource Summary File (RSF)

Field	Description	Format	Comments
1- 6	Form Number	A(6)	Eg. C00633
7- 8	Sequence Number	I(2)	
9- 10	Project Code	I(2)	
11	Resource Type	A(1)	M = Manpower C = Computer O = Other (Services)
12- 16	Resource Code	I(5)	From Encoding Dictionary (Programmer code, computer code, or services code)
17- 22	Form Date	I(6)	YYMMDD
23- 25	Percent Management	I(3)	
26- 31	Beginning Date of Data	I(6)	YYMMDD
32-108	Resources Used	11I(3),F(4.1)	Up to 11 alternating fields representing number of computer runs, number of hours in tenths of hours
109	Status Flag	A(1)	1 = Unchecked from GFSC 2 = Unchecked from UM 5 = Hand checked
110	Phase Flag	A(1)	R = Requirements D = Development M = Maintenance

APPENDIX D

RECORD FORMATS for the V&V DATASET

Record Format
Independent Verification and Validation Dataset
Project Data Record

Field	Description	Format	Comments
1- 2	Project Code	A(2)	
3- 4	Project Application	A(1)	C = CCCI O = Other
6	Type of Development Effort	A(1)	I = Initial Development M = Modification
8	Operating Mode	A(1)	R = Real-time N = Non real-time B = Both
10	Security Classification	A(1)	C = Classified U = Unclassified P = Partial Classification
12	Language Type	A(1)	H = High order language A = Assembly language B = Both
14- 15	HOL Source Lines	I(2)	In thousands
17- 18	Assembly Source Lines	I(2)	In thousands
20	Programming Practices	A(1)	T = Traditional M = Modern
22	Use of Top Down Design at System Level	A(1)	Y = Yes N = No X = Not applicable
24	Use of Top Down Design at Program Level	A(1)	Y = Yes N = No
26	Type of Program Support Library Used	A(1)	F = Full B = Basic M = Manual N = None
28	Type of Structured Code Used	A(1)	S = Simulated Constructs P = Preprocessor D = Directly Compilable N = Not Used
30	Type of RADC Standards Used	A(1)	S = Standard C = With Code reading W = With Design/Code reviews N = Not used
32	Type of Programmer Team Used	A(1)	F = Full programmer team M = Modified programmer team N = Not used
34- 37	Development Start Date	I(4)	MMYY
39- 42	End Requirements Definition Phase	I(4)	MMYY
44- 47	Start Design Phase	I(4)	MMYY
49- 52	End Design Phase	I(4)	MMYY
54- 57	Start Coding Phase	I(4)	MMYY
59- 62	End Coding Phase	I(4)	MMYY
64- 67	Start Testing Phase	I(4)	MMYY
69- 72	Development End Date	I(4)	MMYY

Record Format
Independent Verification and Validation Dataset
Project Data Record continued

Field	Description	Format	Comments
74	Number of Errors Reported in Operations	I(1)	
76- 78	Reason for Maintenance if Performed	A(3)	Up to 3 responses R = Requirements Change E = Error Correction O = Other N = None Performad
80	Contracter / Developer Relationship	A(1)	G = Good F = Fair P = Poor

Record Format
Independent Verification and Validation Dataset
Anomaly Data Record (ADR)

Field	Description	Format	Comments
1- 2	Project Code	A(2)	
4- 7	Anomaly Report Number	A(4)	Left justified
9- 12	Report Part	A(4)	Left justified
14- 19	Report Date	I(6)	MMDDYY
21- 26	Analyst Codes	3I(2)	Up to 3 responses
28- 29	Anomaly Location	A(2)	S = System/Segment Specs. I = Interface Specs. R = Requirements Specs. D1 = Pre-code Design Specs. C = Code D2 = Post-code Design Specs. U = User documentation X = Other
32- 33	Anomaly Category	A(2)	If Location = S,I,or R 1 = Incorrect Requirements 2 = Inconsistent Reqs. 3 = Incomplete Requirements 4A = Unclear Requirements 4B = Unfeasible Reqs. 4C = Extraneous Reqs. 5A = Standards not used 5B = Configuration mngt. 5C = Other presentation prob. If Location = D1 or C 1 = Req./Design compliance 2 = Choice of algorithm 3 = Sequence of operations 4 = Data definition 5A = Initialization 5B = Addressing, indexing 5C = Flags 5D = Counters 5E = Shared memory locations 5F = Other data prob. 6 = Timing / Interruptibility 7A = Input handling 7B = Output 7C = Hardware interface 7D = External SW interface 7E = Routine interface 8A = Extraneous Design/Code 8B = Program error handling 8C = Other design/code prob. 9A = Standards not used 9B = Configuration mngt. 9C = Comments, Annts. 9D = Other presentation prob.

Record Format
Independent Verification and Validation Dataset
Anomaly Data Record continued

Field	Description	Format	Comments
32- 33	Anomaly Category continued	A(2)	If Location = D2 or U 1 = Incorrect documentation 2 = Inconsistent documentation 3 = Incomplete documentation 4 = Other content prob. 5A = Standards not used 5B = Configuration mngt. 5C = Other presentation prob. If Locaion = X 1 = Hardware System prob. 2 = Other Documentation 3 = Unknown origin 4 = Development process 5 = Other Prob.
35- 37	Special Circumstances	3A(1)	Up to 3 responses P = Error from correction of previosuly reported error D = Disagreement among materials with none clearly wrong N = Non optimal decision L = Latent error H = Holdover from previously IV&V effort C = Copy of previous anomaly X = Other
39- 42	Anomaly Effect	4A(1)	Up to 4 responses D = Development V = Verifiability O = Operations M = Maintainability E = Ease of use X = Other
44- 46	Operational Effect	3A(1)	Up to 3 responses C = Correctness A = Accuracy / precision S = Security E = Efficiency X = Other
48	Anomaly Severity	A(1)	H = High M = Medium L = Low U = Unknown
50- 51	IV&V Phase when Detected	2A(1)	Up to 2 responses R = Requirements Verification D = Design Verification C = Code Verification T = Testing Verification W = Documentation Verification X = Other Blank = Unknown

Record Format
Independent Verification and Validation Dataset
Anomaly Data Record continued

Field	Description	Format	Comments
53	Development Phase when Detected	A(1)	R = Requirements D = Design C = Coding and Checkout T = Testing P = Post-testing O = Operations X = Other Blank = Unknown
55- 57	Methods Used	3A(1)	Up to 3 responses M = Manual Analysis E = Execution testing T = Tool use Blank = Unknown
59- 64	Tools Used	2A(3)	Up to 2 responses SIM = Simulation RTA = Real-time Analyzer TED = Text editor ERA = Extension register analyzer MAP = Memory analysis program ICS = Interpretative computer simulator MSS = Missile system simulation CMP = Compiler ASM = Assembler Blank = Unknown or none
66	Anomaly Acceptance	A(1)	A = Accepted as written C = accepted with changes R = Rejected W = Withdrawn or superceded U = Unknown X = Other
68- 69	Action Taken	A(2)	F = Fixed and verified FW = Fixed found wrong FU = Fixed, unchecked C = Negated by unrelated change D = Deferred DF = Deferred,fixed later DN = Deferred, not fixed W = Work around used N = No action taken O = Still open U = Unkown P = Partially fixed X = Other

Record Format
Independent Verification and Validation Dataset
Anomaly Data Record continued

Field	Description	Format	Comments
71- 73	Materials Changed	A(3)	Up to 3 responses S = System/subsystem spec. I = Interface Spec. R = Requirements Spec. D = Design Spec. C = Code U = User documentation X = Other Blank = Unknown
75- 80	Resolution Date	I(6)	MMDDYY

APPENDIX E

RECORD FORMATS for the ARF ERROR DATASET

Project level data is recorded in the same format as data contained in the DACS Productivity Dataset. See Appendix A of this report for these formats.

Error Report data is recorded in the same format as Chage Report data recorded in the NASA/SWL dataset. See the record formats for the Change Report File in Appendix C of this report for this information

Segment or Module Descriptive Data is recorded in the following formats:

Location	Parameter	Format	Comments
1 - 2	Project Code	I(2)	
3 - 5	Component Code	I(3)	001-253
6 - 9	Total Statements in Segment	I(4)	
10 - 12	Number of Comments in Segment	I(3)	
13 - 15	Number of Pre-Processor Statements in Segment	I(3)	
16	Subjective Complexity	A(1)	E - Easy M - Moderate H - Hard

APPENDIX F

RECORD FORMATS for the O&M DATASET

Record Format
PAVE PAWS Operations and Maintenance Dataset
Programmer Experience Profile (PEP) Record 1

Field	Description	Format	Comments
1- 2	Record Type	A(2)	"1A"
3- 7	Personnel Identification Code	A(5)	Uniquely Assigned
8-27	Filler		
28-29	Age	I(2)	
30-35	Date Profile Completed	I(6)	YYMMDD
36-50	Project Name	A(15)	
51-70	Job Title	A(20)	
71-75	Position	A(4)	Eg. GS12
76-78	Section Name	A(3)	Eg. ADQ
79	Years of College Education	I(1)	
80	Years of High School Education	I(1)	

Record Format
PAVE PAWS Operations and Maintenance Dataset
Programmer Experience Profile (PEP) Record 2

Field	Description	Format	Comments
1- 2	Record Type	A(2)	"1B"
3- 4	Year Graduated	I(2)	YY (High School if no College)
5- 7	First College Degree	A(3)	Eg. BS, AAS
8- 9	Degree Year	I(2)	YY
10-13	Major Course of Study	A(4)	
14-16	Second College Degree	A(3)	Eg. MS, MEE
17-18	Degree Year	I(2)	YY
19-22	Major Course of Study	A(4)	
23-25	Third College Degree	A(3)	Eg. PHD, MAT
26-27	Degree Year	I(2)	YY
28-31	Major Course of Study	A(4)	
32-33	Number Computer Science Courses	I(2)	
34-35	Number Computer Science Semester Hours	I(2)	
36-37	Number Computer Science Seminars	I(2)	
38-39	Number Years with Computers	I(2)	
40-41	Percent of Years Individual Effort	I(2)	
42-43	Percent of Years Team Effort	I(2)	
44-45	Percent of Years Supervisor	I(2)	
46-47	Number Years w/Structured Programming	I(2)	To nearest year
48-49	Number Years w/PDL	I(2)	To nearest year
50-51	Number Years w/HIPO Charts	I(2)	To nearest year
52-53	Number Years w/Top Down Development	I(2)	To nearest year
54-55	Number Years w/PSL	I(2)	To nearest year
56-57	Number Years w/Precompilers	I(2)	To nearest year
58-59	Number Years w/Chief Programmer Team	I(2)	To nearest year
60-61	Number Years w/Other Techniques	I(2)	To nearest year
62-63	Years Experience with JOVIAL	I(2)	To nearest year
64-65	Years Experience with Assembler	I(2)	To nearest year
66-67	Years Experience with FORTRAN	I(2)	To nearest year
68-69	Years Experience with COBOL	I(2)	To nearest year
70-71	Years Experience with ALGOL	I(2)	To nearest year
72-73	Years Experience with PL/1	I(2)	To nearest year
74-75	Years Experience with PASCAL	I(2)	To nearest year
76-77	Years Experience with Other Languages	I(2)	To nearest year

Record Format
PAVE PAWS Operations and Maintenance Dataset
Programmer Experience Profile File (PEP) Record 3

Field	Description	Format	Comments
1- 2	Record Type	A(2)	"1C"
3- 4	Number Years w/Business Appl.	I(2)	To nearest year
5- 6	Number Years w/Scientific/Math Appl.	I(2)	To nearest year
7- 8	Number Years w/Sys. Prog. Appl.	I(2)	To nearest year
9-10	Number Years w/Real-time Appl.	I(2)	To nearest year
11-12	Number Years w/Database Appl.	I(2)	To nearest year
13-14	Number Years w/Other Appl.	I(2)	To nearest year
15-21	Name of Primary Computer	A(7)	Eg. IBM360
22-28	Name of Primary Operating System	A(7)	Eg. DOS
29-30	Experience on Primary System	I(2)	To nearest year
31-37	Name of Secondary Computer	A(7)	
38-44	Name of Secondary Operating System	A(7)	
45-46	Experience on Secondary System	I(2)	To nearest year
47-53	Name of Tertiary Computer	A(7)	
54-60	Name of Tertiary Operating System	A(7)	
61-62	Experience on Tertiary System	I(2)	To nearest year

Record Format
PAVE PAWS Operations and Maintenance Dataset
Discrepancy Report History File (DRH) Record 1

Field	Description	Format	Comments
1- 2	Record Type	A(2)	"2A"
3- 7	Discrepancy Report (DR) Number	A(5)	YNNNN Y = Last Digit of Year NNNN = Sequence Number
9-11	DR Origin	A(3)	Eg. 7th
13-16	Local DR Number	A(4)	Eg. B270
18-57	DR Description	A(40)	
59-65	Date DR Received	A(7)	Eg. 10 Oct 79
67-69	Responsible Section	A(3)	Eg. TAC
71	DR Priority	A(1)	R = Routine U = Urgent E = Emergency

Record Format
PAVE PAWS Operations and Maintenance Dataset
Discrepancy Report History File (DRH) Record 2

Field	Description	Format	Comments
1- 2	Record Type	A(2)	"2B"
3- 9	Date Analysis Started	A(7)	Eg. 06 Nov 79
11-17	PMR/PCD/PDDR Open Date	A(7)	Eg. 12 Dec 79
19-57	Remarks	A(40)	

Record Format
PAVE PAWS Operations and Maintenance Dataset
Discrepancy Report History File (DRH) Record 3

Field	Description	Format	Comments
1- 2	Record Type	A(2)	"2C"
3-17	Filler		
19-25	Estimated Completion Date	A(7)	Eg. 10 Feb 80
27-33	Scheduled Version Release	A(7)	Eg. PTAC-DO
39-41	Scheduled Operation Date	A(7)	Eg. 12 Apr 80
43-49	Date Forwarded to NCCB	A(7)	Eg. 18 Dec 79
51-53	NCCB Action	A(3)	APP = Approved REJ = Rejected DEF = Deferred
55-61	Date of NCCB Action	A(7)	Eg. 31 Dec 79
63-69	Date DR Closed	A(7)	Eg. 13 Apr 80

Record Format
PAVE PAWS Operations and Maintenance Dataset
Maintenance Activity File (MAF) Record 1

Field	Description	Format	Comments
1- 5	Discrepancy Report (DR) Number	I(5)	
6-11	Date DR Submitted	I(6)	YYMMDD
12-15	CPCI Affected	A(4)	
16-19	CPCG Affected	A(4)	
20-24	CPC Affected	A(5)	Blank if more than one
25-26	Primary Maintenance Activity	A(2)	E1 = Error correction A1 = Add capability D1 = Delete capability O1 = Optimize/enhance
27-28	Secondary Maitnenance Activity	A(2)	E2 = Error Correction A2 = Add capability D2 = Delete capability O2 = Optimize/enhance Blank = not needed
29-30	Version Release Affected	A(2)	
31-32	Precision of Change Specification	A(2)	VP = Very Precise P = Precise IM = Imprecise
33	Urgency Code	A(1)	E = Emergency U = Urgent R = Routine
34-35	Complexity of Change	A(2)	VC = Very complex C = Complex M = Medium Complexity S = Simple VS = Very Simple
36-37	Primary Means of Detection	A(2)	1 Response each for two fields HP = Hand processing
38-39	Secondary Means of Detection	A(2)	PC = Personal communication IL = Infinite loop MC = Maintenance Crosscheck IE = Interrupt error IO = Incorrect output MO = Missing output EM = Error message CR = Code review DR = Documentation review SD = Special debug code OT = Other
40-46	Programmer Code	A(7)	
47-50	Manhours to Implement Change	I(4)	
51-54	Computer Hours for Change	F(4.1)	Tenths of CPU Hours
55-57	Number of Segments Affected	I(3)	Specifies number of Type 2 records following
58-63	Sources of Error	3A(2)	Up to three responses MS = Specs. Misinterpreted IS = Specs. Incorrect NS = Specs. Incomplete

Record Format
PAVE PAWS Operations and Maintenance Dataset
Maintenance Activity File (MAF) Record 1 continued

Field	Description	Format	Comments
58-63	Sources of Error continued		SF = Function not Implemented SI = Interface not Implemented SO = Software Interface HI = HW/SW Interface OS = Operating System LE = Logic Error CE = Computational Error DE = Data I/O Error DD - Data Definition Error CN = Cause Not Found IO = I/O software PM = Prior Modification SS = Support Software DS = Deck Setup Error OE = Operator Error OT = Other
64-69	Nature of Change	3A(2)	Up to 3 responses DO = Documentation FI = Fix Instruction CC = Change Constants ST = Structural Change AL = Algorithmic OT = Other
70-75	Type of New Requirement	3A(2)	Up to 3 responses MI = Mission Changed EM = New Engineering Model SW = More Efficient Software HW = Change in Hardware SS = New Support Software OT = Other

Record Format
PAVE PAWS Operations and Maintenance Dataset
Maintenance Activity File (MAF) Record 2

Field	Description	Format	Comments
1- 5	Discrepancy Report Number	I(5)	
6	Segment Status	A(1)	A = Segment being added D = Segment being deleted C = Segment being changed
7-46	Name of Segment Affected	A(35)	

Record Format
PAVE PAWS Operations and Maintenance Dataset
CPCG Description File (CDF)

Field	Description	Format	Comments
1- 4	CPCI Name	A(4)	
5-13	CPCG Name	A(9)	
14	Special Display constraint	A(1)	Y = Yes N = No
15	Detailed Requirements Definition	A(1)	Y = Yes N = No
16	Change to Operational Requriements	A(1)	Y = Yes N = No
17	Real Time Operation	A(1)	Y = Yes N = No
18	CPU Memory Constraint	A(1)	Y = Yes N = No
19	CPU Time Constraint	A(1)	Y = Yes N = No
20	First Software Developed on CPU	A(1)	Y = Yes N = No
21	Developed Concurrently with HW	A(1)	Y = Yes N = No
22	Time Sharing (vs Batch)	A(1)	Y = Time Sharing N = Batch
23	Developer used other Facility	A(1)	Y = Yes N = No
24	Operational Site Development	A(1)	Y = Yes N = No
25	Not Developed on Target System	A(1)	Y = Yes N = No
26	Programmer Access to Computer	A(1)	Y = Direct Access N = Indirect Access
27-32	PSL Management Report Date	I(6)	YYMMDD
33-35	PSL Library Level	A(3)	
36-38	Source of Data	A(3)	SEG = PSL Summary by Segment PRG = PSL Summary by Program
39-41	Number of Programs	I(3)	
42-45	Number of Segments	I(4)	
46-51	Number of Source Lines	I(6)	
52-57	Words of Object Code	I(6)	

Record Format
PAVE PAWS Operations and Maintenance Dataset
Program Change History File (PCH)

Field	Description	Format	Comments
1-40	Program Long name	A(40)	
42-47	Program Short name	A(6)	
49-52	Language	A(4)	
55-61	Date Program was Last Changed	A(8)	YY/MM/DD
64-71	Time Program was Last Changed	A(8)	HH.MM.SS
72-76	Number of Segments in Program	I(5)	
77-82	Total Size	I(6)	In source lines
83-86	Number of Stubs	I(4)	
88-89	Program Version	A(2)	Maximum of all segment versions
90-93	Program Edition	I(4)	Sum of segment aditions
94-98	Program Instance	I(5)	Incremental for each compile
101-108	Date Compiled	A(8)	YY/MM/DD
110-117	Time Compiled	A(8)	HH.MM.SS
118-123	Object Size	I(6)	In decimal words

Record Format
PAVE PAWS Operations and Maintenance Dataset
CPCG Status File (CSF)

Field	Description	Format	Comments
1- 4	CPCG Name	A(4)	
5-10	PSL Management Report Date	I(6)	YYMMDD
11-15	Effective Code at PSL PRG Level	I(5)	In source lines
16-17	Highest Version at PRG Level	A(2)	
18-22	Effective Code at PSL CPT Level	I(5)	In source lines
23-24	Highest Version at CPT Level	A(2)	
25-29	Effective Code at PSL INT Level	I(5)	In source lines
30-31	Highest Version at INT Level	A(2)	
32-36	Effective Code at PSL FIX Level	I(5)	In source lines
37-38	Highest Version at FIX Level	A(2)	
39-43	Effective Code at PSL TST Level	I(5)	In source lines
44-45	Highest Version at TST Level	A(2)	
46-50	Effective Code at PSL FRZ Level	I(5)	In source lines
51-52	Highest Version at FRZ Level	A(2)	
53-57	Effective Code at PSL DEL Level	I(5)	In source lines
58-59	Highest Version at DEL Level	A(2)	
60-64	Durable Code at PSL PRG Level	I(5)	In source lines
65-66	Highest Version at PRG Level	A(2)	
67-71	Durable Code at PSL CPT Level	I(5)	In source lines
72-73	Highest Version at CPT Level	A(2)	
74-78	Durable Code at PSL INT Level	I(5)	In source lines
79-80	Highest Version at INT Level	A(2)	
81-85	Durable Code at PSL FIX Level	I(5)	In source lines
86-87	Highest Version at FIX Level	A(2)	
88-92	Durable Code at PSL TST Level	I(5)	In source lines
93-94	Highest Version at TST Level	A(2)	
95-99	Durable Code at PSL FRZ Level	I(5)	In source lines
100-101	Highest Version at FRZ Level	A(2)	
102-106	Durable Code at PSL DEL Level	I(5)	In source lines
107-108	Highest Version at DEL Level	A(2)	
109-114	Last Change at PRG Level	I(6)	YYMMDD
115-120	Last Change at CPT Level	I(6)	YYMMDD
121-126	Last Change at INT Level	I(6)	YYMMDD
127-132	Last Change at FIX Level	I(6)	YYMMDD
133-138	Last Change at TST Level	I(6)	YYMMDD
139-144	Last Change at FRZ Level	I(6)	YYMMDD
145-150	Last Change at DEL Level	I(6)	YYMMDD

Record Format
PAVE PAWS Operations and Maintenance Dataset
Segment Change History File (SCH)

Field	Description	Format	Comments
1-40	Segment Longname	A(40)	
42-47	Segment Shortname	A(6)	
49-52	Language	A(4)	
54-57	Segment Type	A(4)	
59-66	Date Segment was Created	A(8)	YY/MM/DD
67-70	Current Source Lines in Segment	I(4)	
71-74	Gross Size of Segment	I(4)	includes deleted lines
76-83	Date Segment Last Changed	A(8)	YY/MM/DD
85-92	Time Segment Last Changed	A(8)	HH.MM.SS
94-95	Segment Version	A(2)	
96-99	Segment Edition	I(4)	
100-103	Total Changes to Segment	I(4)	
104-107	Changes to Current Segment Version	I(4)	
108-111	Gross Size of Current Version	I(4)	includes deleted lines
113-118	Identification of Segment Creater	A(6)	
120	Special Circumstances Flag	A(1)	
123-130	Identification of Person Last Changing Segment	A(6)	

APPENDIX G

DATA-ITEMS CONTAINED in the BSDS DATASET

GLOSSARY OF DATA-ITEM NAMES IN THE BSDS DATASET

PROJ-ID	PROJECT IDENTIFICATION
PROJ-VERSION	PROJECT VERSION
PROJ-TYPE	PROJECT TYPE
SYS-ID	SYSTEM IDENTIFICATION
SYS-VERSION	SYSTEM VERSION
SYS-TYPE	SYSTEM TYPE
SSYS-ID	SUBSYSTEM OR FUNCTIONAL AREA IDENTIFICATION
SSYS-VERSION	SUBSYSTEM VERSION
SSYS-TYPE	SUBSYSTEM TYPE
MOD-ID	MODULE IDENTIFICATION
MOD-VERSION	MODULE VERSION
MOD-TYPE	MODULE TYPE
COMP-ID	COMPUTER IDENTIFICATION
COMP-OM	COMPUTER OPERATING MODE
COMP-RATE	COMPUTER PROCESSING RATE
COMP-OS	COMPUTER OPERATING SYSTEM TYPE
TECH-ID	IDENTIFICATION OF THE CONSTRUCTION TECHNOLOGY
COMPL-ID	TYPE OF COMPLEXITY MEASURE USED
COMPLEXITY	THE COMPLEXITY MEASURE VALUE
CONST-TYPE	CONSTITUENT TYPE(EX. JOVIAL,ASSEMBLY LANGUAGE)
NUM-OCCUR	NUMBER OF OCCURRENCES OF CONSTITUENT TYPE
PHASE	PHASE IN WHICH ACTION OCCURRED
NUM-RUNS-TOT	TOTAL NUMBER OF RUNS
TEST-PER	THE PERIOD IN WHICH THE TEST WAS PERFORMED
NUM-RUNS-OK	TOTAL NUMBER OF CORRECT RUNS
AHRS-PER-TEST	AVERAGE NUMBER OF HOURS PER TEST
TEST-ID	TEST IDENTIFICATION
TEST-TYPE	TYPE OF TEST
DATE-RUN	DATE THE TEST WAS RUN
STRESS-TYPE	TYPE OF STRESS APPLIED
STRESS-MEAS	AMOUNT OF STRESS APPLIED
TEST-RESULT	RESULT OF TEST

GLOSSARY OF DATA-ITEM NAMES IN THE BSDS DATASET

NUM-ERR	NUMBER OF ERRORS DISCOVERED PER TEST
SPR-NUM	SOFTWARE PROBLEM REPORT NUMBER
DATE-OPEN	DATE THE PROBLEM WAS REPORTED
MOD-SOURCE	THE MODULE ID WHERE THE PROBLEM WAS MANIFESTED
ERR-CAT-TYPE	ERROR CATEGORY TYPE
ERROR-CAT	ERROR CATEGORY CODE
SEV-TYPE	SEVERITY TYPE
SEVERITY	SEVERITY
TYPE-TERM	TYPE OF TERMINATION
HRS-TO-DISC	HOURS TO DISCOVERY
WORK-CAT	THE TYPE OF DEVELOPMENT TASK PERFORMED
SMN-NUM	SOFTWARE MODIFICATION NOTICE NUMBER
MOD-CHANGED	THE ID OF THE CHANGED MODULE
MOD-CH-VERS	THE VERSION OF THE CHANGED MODULE
COR-TYPE	CORRECTION TYPE
COR-MECH	CORRECTION MECHANISM
ACT-CAT	THE TYPE OF TEST PERFORMED
DATE-BEGUN	DATE WHEN PROBLEM SOLUTION WAS INITIATED
DATE-CLOSE	DATE WHEN PROBLEM WAS REPORTED TO BE CLOSED
DAYS-OPEN	NUMBER OF DAYS BETWEEN DATE OPEN AND DATE CLOSE
HHRS-TO-FIX	HUNDRETHS OF HOURS TO FIX
NUM-CHANGED	NUMBER OF SOURCE STATEMENTS CHANGED
CODE-CONT	A CODE THAT INDICATES AN SPR DOCUMENTS MORE THAN 1 PROBLEM
PROB-DESC	A DESCRIPTION OF THE PROBLEM
CORR-DESC	A DESCRIPTION OF THE CORRECTION
ERROR-DESC	A DESCRIPTION OF THE ERROR

SOURCE: [DUVALL-79]

APPENDIX H

RECORD FORMAT for the
DACS COMPOSITE PRODUCTIVITY DATASET

Record Format
Composite Productivity Dataset

Field	Description	Format	Comments
1- 4	Project Code	I(4)	
5- 11	Project Size	I(7)	In source lines
12- 16	Project Development Effort	I(5)	In manmonths
17- 19	Project Schedule Duration	I(3)	In months
20- 22	Average Staff Size	I(3)	
23- 27	Productivity	I(5)	Lines per Manmonth
28- 32	Number of Documented Changes	I(5)	
33- 35	Program Design Language Usage	I(3)	Expressed as percent 999 = Data no recorded
36- 38	Structured Code Usage	I(3)	Expressed as percent 999 = Data not recorded
39- 41	Top Down Programming Usage	I(3)	Expressed as percent 999 = Data not recorded
42- 44	Chief Programmer Team Usage	I(3)	Expressed as percent 999 = Data not recorded
45- 47	Programmer / Librarian Usage	I(3)	Expressed as percent 999 = Data not recorded
48- 50	Formal Code Review Usage	I(3)	Expressed as percent 999 = Data not recorded
51- 61	Primary Language	A(11)	
62	Dataset Origin Code	A(1)	R = DACS Productivity Dataset N = NASA-SEL Software Eng. Dataset V = IV&V Dataset